AMERICAN SOCIETY FOR MICROBIOLOGY | Journal of Virology®

GENETIC DIVERSITY AND EVOLUTION

# Phylodynamics of HIV in the Mexico City Metropolitan Region

**Santiago Avila-Rios,[a] Claudia García-Morales,[a] Gustavo Reyes-Terán,[b] Andrea González-Rodríguez,[c] Margarita Matías-Florentino,[a] Sanjay R. Mehta,[d,e] Antoine Chaillon[d]**

[a]Center for Research in Infectious Diseases, National Institute of Respiratory Diseases, Mexico City, Mexico
[b]Coordinating Commission of the National Institutes of Health and High Specialty Hospitals, Ministry of Health, Mexico City, Mexico
[c]Clínica Especializada Condesa, Mexico City, Mexico
[d]Division of Infectious Diseases and Global Public Health, University of California, San Diego, San Diego, California, USA
[e]Veterans Affairs Health System, San Diego, California, USA

Sanjay R. Mehta and Antoine Chaillon contributed equally to this article.

**ABSTRACT** Evolutionary analyses of viral sequences can provide insights into transmission dynamics, which in turn can optimize prevention interventions. Here, we characterized the dynamics of HIV transmission within the Mexico City metropolitan area. HIV *pol* sequences from persons recently diagnosed at the largest HIV clinic in Mexico City (between 2016 and 2021) were annotated with demographic/geographic metadata. A multistep phylogenetic approach was applied to identify putative transmission clades. A data set of publicly available sequences was used to assess international introductions. Clades were analyzed with a discrete phylogeographic model to evaluate the timing and intensity of HIV introductions and transmission dynamics among municipalities in the region. A total of 6,802 sequences across 96 municipalities (5,192 from Mexico City and 1,610 from the neighboring State of Mexico) were included (93.6% cisgender men, 5.0% cisgender women, and 1.3% transgender women); 3,971 of these sequences formed 1,206 clusters, involving 78 municipalities, including 89 clusters of $\geq$10 sequences. Discrete phylogeographic analysis revealed (i) 1,032 viral introductions into the region, over one-half of which were from the United States, and (ii) 354 migration events between municipalities with high support (adjusted Bayes factor of $\geq$3). The most frequent viral migrations occurred between northern municipalities within Mexico City, i.e., Cuauhtémoc to Iztapalapa (5.2% of events), Iztapalapa to Gustavo A. Madero (5.4%), and Gustavo A. Madero to Cuauhtémoc (6.5%). Our analysis illustrates the complexity of HIV transmission within the Mexico City metropolitan area but also identifies a spatially active transmission area involving a few municipalities in the north of the city, where targeted interventions could have a more pronounced effect on the entire regional epidemic.

**IMPORTANCE** Phylogeographic investigation of the Mexico City HIV epidemic illustrates the complexity of HIV transmission in the region. An active transmission area involving a few municipalities in the north of the city, with transmission links throughout the region, is identified and could be a location where targeted interventions could have a more pronounced effect on the entire regional epidemic, compared with those dispersed in other manners.

**KEYWORDS** HIV, dispersal, Bayesian discrete phylogeography, Mexico, phylodynamics

HIV remains an important transmissible disease worldwide, contributing to significant morbidity and deaths each year (1). HIV sequences collected for drug resistance testing as part of routine care for persons with HIV (PWH) can be used to infer the dynamics of transmission using phylogenetic approaches. If geospatial and temporal data associated with each sequence are incorporated into these analyses, then the spatiotemporal spread of HIV in a region can be deduced. Further linking of clinical and

**Editor** Guido Silvestri, Emory University

Address correspondence to Antoine Chaillon, achaillon@health.ucsd.edu.
The authors declare no conflict of interest.

**Received** 4 May 2022
**Accepted** 6 June 2022
**Published** 28 June 2022

demographic data can provide a detailed understanding of the spread of the virus across a region and provide insights into the factors that may be driving or impacting viral transmission.

Phylodynamic and phylogeographic analyses can be used not only to describe epidemiology but also to guide the choice and implementation of interventions designed to prevent further HIV transmission, by highlighting key high-risk populations and regions where limited prevention efforts may have an impact that significantly extends to the entire region (2–7). A detailed understanding of the underlying transmission dynamics in a region could also provide the ability to monitor an epidemic for sudden changes, evaluate the impact of public health interventions, or modify public policy. However, as for all molecular epidemiologic studies, these inferences and conclusions are impacted by the sampling depth (i.e., ratio of the number of samples to the number of PWH in a geospatial unit) and representativity (i.e., sampling by demographic factors) of the sampled population.

Mexico has had a progressive approach to the HIV epidemic since its early stages (8) and has provided free antiretroviral therapy (ART) to all PWH since 2003 through the Centros Ambulatorios para la Prevención y Atención de VIH/SIDA e Infections de Transmisión Sexual (CAPASITS), the Servicios de Atención Integral (SAIH) network of clinics for uninsured persons, and social security clinics. The Condesa Clinic in Mexico City is one of the largest HIV care providers in Mexico. It offers a comprehensive suite of services, including testing and counseling for HIV and other sexually transmitted infections (STIs) and mental health, HIV, and transgender care, with laboratory testing and retention services. In addition to caring for over 19,000 PWH, the clinic has an active screening and prevention program, with over 25,000 persons tested for HIV and 3,500 to 4,000 new HIV infections diagnosed annually (9). Centralization of nearly 70% of the local new diagnoses in the Condesa clinic represents a unique opportunity to perform molecular surveillance and phylogeographic analyses. Since 2016, the Center for Research in Infectious Diseases of the National Institute of Respiratory Diseases (CIENI/INER), a reference laboratory for HIV sequencing, has established a close collaboration with the Condesa clinic to perform baseline HIV genotyping for all recently diagnosed persons and persons returning to care at Condesa for HIV drug resistance, as well as molecular surveillance in the Mexico City metropolitan zone (including Mexico City and neighboring municipalities of the State of Mexico). Notably, Mexico City is also one the most populated metropolitan areas in the world, with close links to the United States, as well as many countries in Central and South America and Europe. Here, we use the data from this impressive program, with centralized screening and treatment, to describe in detail the phylodynamics of HIV in the Mexico City metropolitan zone, focusing on how the region is impacted by the larger national and worldwide epidemics and how the movement of virus within the region impacts the local epidemiology.

## RESULTS

**Overview of data.** A total of 6,802 sequences (5,192 sequences from Mexico City and 1,610 from the neighboring State of Mexico, with 98.6% being HIV-1 subtype B) were included in the study (Fig. 1); 93.6% belonged to cisgender men, 5.0% to cisgender women, and 1.3% to transgender women. Sequences were obtained from all 16 boroughs in Mexico City and 73 (60.3%) of the 121 municipalities in the State of Mexico (Fig. 1). Numbers of new sequences varied with time, and included periods in late 2018, early 2019, and early 2020 when markedly fewer sequences were collected (see Fig. S1 in the supplemental material). See Table 1 for a more complete description of these sequence data.

**Clade identification.** After implementation of the method published by Cuypers et al. (10), 3,971 of the 6,802 sequences formed a total of 1,206 clades or genetically related clades of sequences. These clades involved 78 boroughs in Mexico City and municipalities in the State of Mexico. The number of clades observed per municipality correlated with the number of sequences per population in that municipality ($R = 0.61$, $P < 0.0001$), and clade size correlated with the number of geospatial units included in the clade, both in 100% male clades and in mixed-

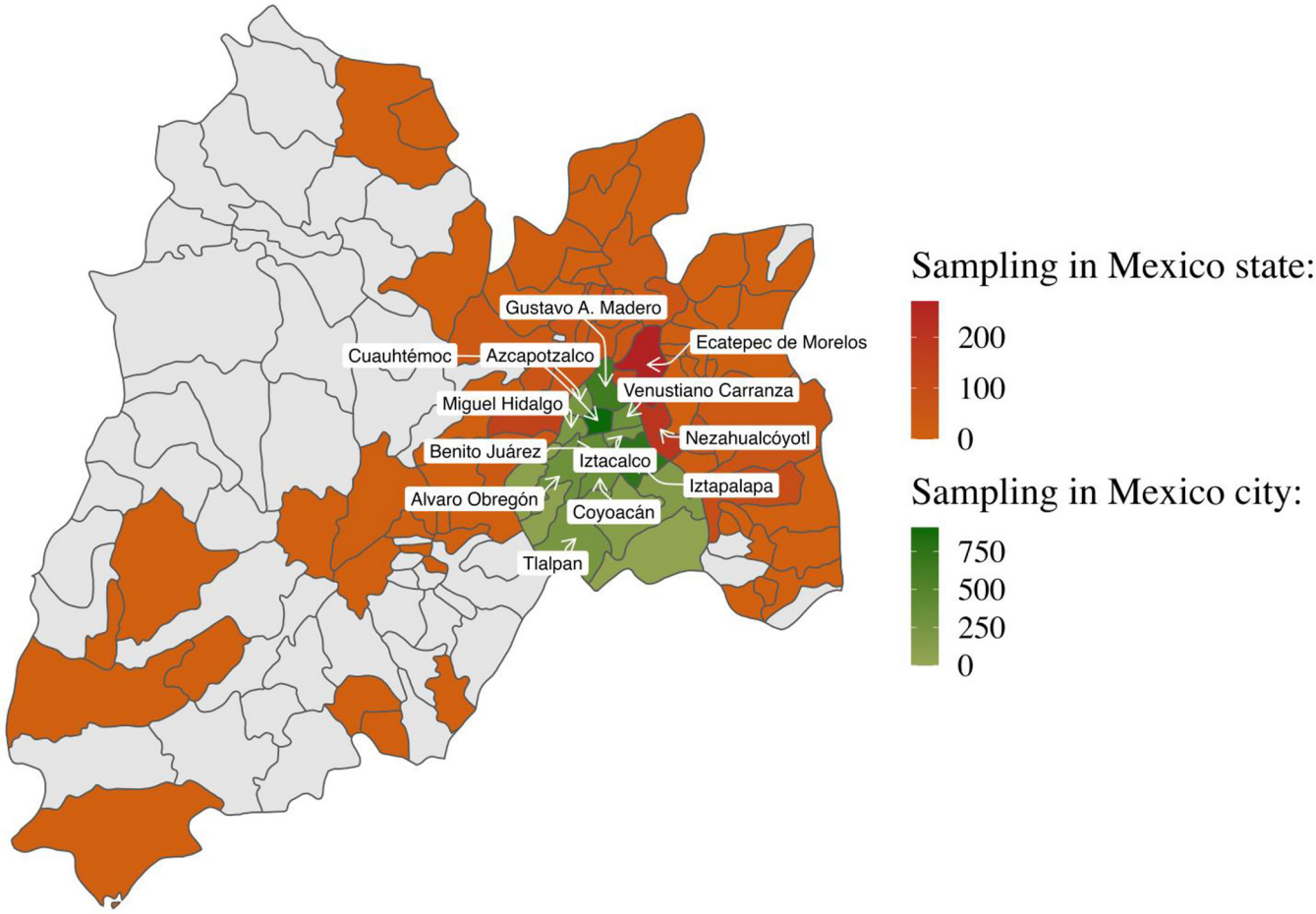


**FIG 1** Sampling distribution. The map shows HIV partial *pol* sequences sampled in the municipalities of Mexico City (green tones) and Mexico State (brown tones). Only municipalities with ≥200 sequences are labeled.

gender clades ($R$ = 0.92 and $R$ = 0.93, respectively, $P < 0.0001$) (see Fig. S2). The majority of the clades identified were limited in size (mean size, 4 sequences [95% confidence interval [CI]), 2 to 5 sequences] (see Fig. S3).

**Viral introductions into Mexico City and the State of Mexico.** Next, we used discrete phylogeographic analysis to first estimate the number of viral introductions into the Mexico City metropolitan region. Of the 1,032 introductions inferred, 394 were from other Mexican states (Fig. 2A), the majority of which were from Quintana Roo ($n$ = 163). Of the 638 viral introductions from international locations, 568 were from the United States, and 75% of those were from California (Fig. 2B). We next estimated the timing of viral introductions into the region and found sustained transmissions into the region, with a peak of viral introductions into the region occurring in 2016 (Fig. 2C).

**Viral migration within Mexico City and the State of Mexico.** As noted in Materials and Methods, to analyze viral migration in the study region we limited our analysis to only clades of ≥10 sequences, to allow for computational feasibility. Among all clades, 89 clades of ≥10 sequences (total of 1,364 sequences, all being HIV-1 subtype B), were identified and analyzed to elucidate transmission dynamics within the Mexico City metropolitan region. To allow simultaneous viewing of sampling date and geographic information, data were uploaded to the MicroReact platform (11) (https://microreact.org/project/223ln8u9trx5cqxdaeyfsu) (see Fig. S4). Discrete Bayesian phylogeographic analysis of these clades revealed 354 transmission events between municipalities and boroughs in the region, with strong support (adjusted Bayes factor [$BF_{adj}$] of ≥3), and a complex viral migration network within the region, with frequent intermixing between municipalities in Mexico City and the State of Mexico (Fig. 3).

**TABLE 1** Cohort description

| Parameter | Data for: Mexico City ($n$ = 5,192) | Data for: State of Mexico ($n$ = 1,610) | $P$ |
|---|---|---|---|
| Gender (no. [%]) | | | 0.014 |
| Female | 269 (5.19) | 72 (4.48) | |
| Male | 4,830 (93.2) | 1,526 (94.9) | |
| Transgender female | 78 (1.51) | 10 (0.62) | |
| Transgender male | 3 (0.06) | 0 (0.00) | |
| Age (mean [95% CI]) (yr) | 30.0 (25.0–36.0) | 27.0 (24.0–33.0) | <0.001 |
| HIV risk (no. [%]) | | | |
| Bisexual | 5 (0.10) | 0 (0.00) | |
| Heterosexual | 431 (8.30) | 108 (6.71) | |
| MSM | 2,246 (43.3) | 755 (46.9) | |
| MSM transgender | 77 (1.48) | 6 (0.37) | |
| Other | 5 (0.10) | 0 (0.00) | |
| PWID | 1 (0.02) | 0 (0.00) | |
| PWID and MSM | 3 (0.06) | 0 (0.00) | |
| Unknown | 2,424 (46.7) | 741 (46.0) | |
| Current $CD4^+$ cell count (mean [95% CI]) (cells/mL) | 213 (105–337) | 191 (95.0–309) | 0.001 |
| HIV RNA level (mean [95% CI]) ($log_{10}$ copies/mL) | 4.66 (4.08–5.19) | 4.64 (4.14–5.14) | 0.978 |

Nevertheless, transitions within Mexico City accounted for the majority of transitions locally, and this pattern was consistent across time (see Fig. S5).

Next, we temporally reconstructed the direction and intensity of viral flows between municipalities in the region (Fig. 3B). The most frequent of the inferred viral migrations occurred between municipalities within Mexico City, i.e., Cuauhtémoc to Iztapalapa (5.2% of events), Iztapalapa to Gustavo A. Madero (5.4%), and Gustavo A. Madero to Cuauhtémoc (6.5%). These data are also shown in a Sankey plot, which more clearly demonstrates the relative contributions of the viral migrations to and from various municipalities (Fig. 4). We also examined the role of risk and time in relation to these viral migration patterns (see Fig. S6A). As expected, the vast majority (86.4%) of viral migration events were associated with transmissions between men who have sex with men (MSM) (see Fig. S6B). Of these events, 68.5% (59.2% of all events) occurred locally between municipalities within Mexico City. Of note, the majority of transmissions involving only heterosexual persons linked municipalities in Mexico City with those in the State of Mexico (77.6% of the transmission events between heterosexual persons, which accounted for 9.63% of all inferred events). We also observed a negative correlation between increasing geospatial distance and the number of transmission events ($R = -0.26$, $P = 0.0025$), with most migration events occurring between municipalities less than 50 km apart (Fig. 5A). Finally, as expected, we also found a positive correlation between the population density (individuals per square mile) of the municipality and the total number of transmission events from/to these locations ($R = 0.47$, $P < 0.001$) (Fig. 5B), with Cuauhtémoc, the most densely populated municipality, being found to be the main hub of transmission events.

## DISCUSSION

This phylogeographic and phylodynamic analysis of the HIV epidemic in the Mexico City metropolitan area highlights a highly dynamic regional epidemic that also demonstrates interactions with most of the rest of the states in Mexico and countries in the Americas, Europe, and Asia. This is expected, given that Mexico City is one of the largest metropolitan areas in the world, central to Mexico, with cultural and commercial links around the world. Interestingly, we found that the state of Quintana Roo (which includes popular tourist destinations such as Cancun, Playa del Carmen, and Tulum) was a major source of viral introductions into the Mexico City epidemic from within

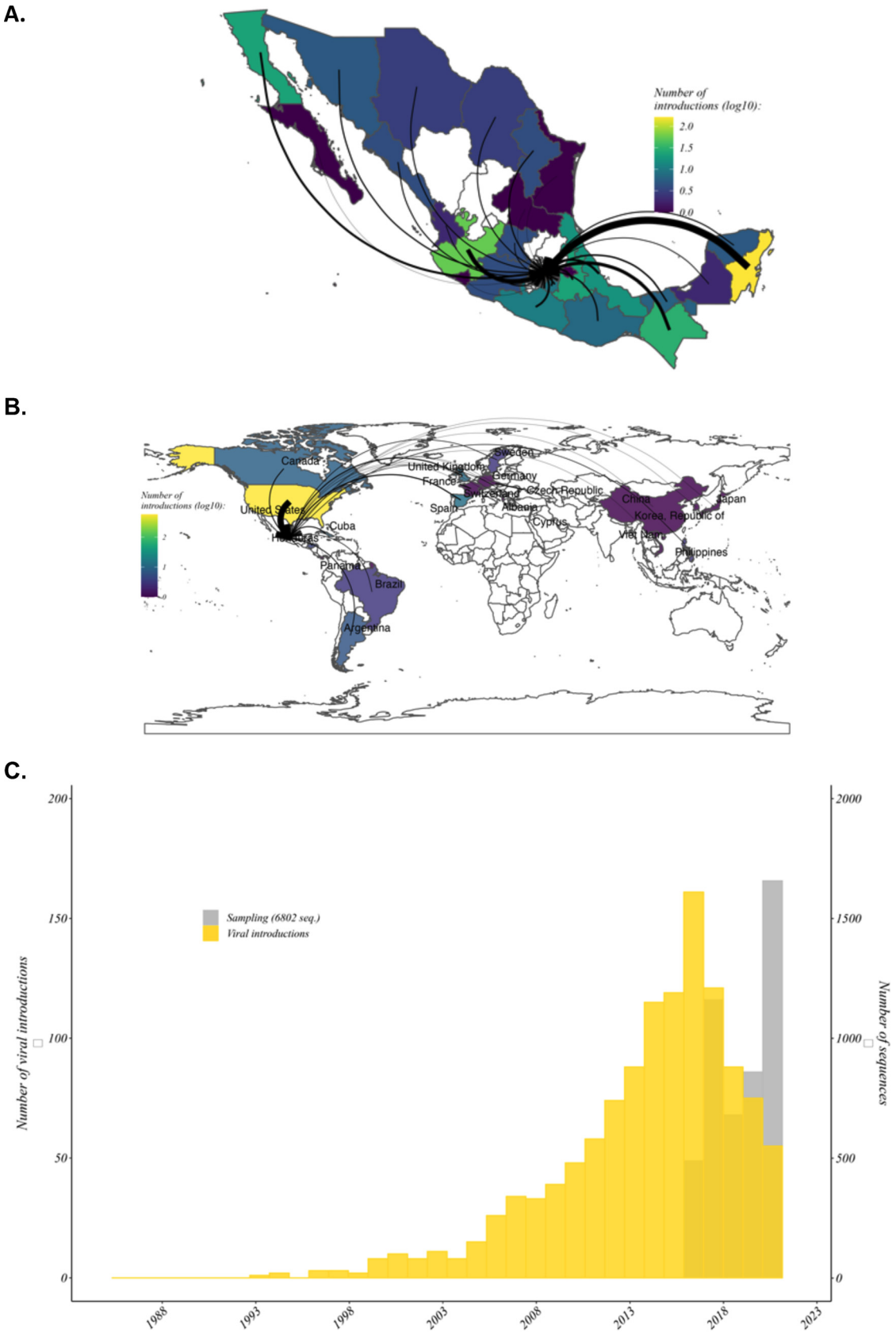


**FIG 2** Viral introductions within the Mexico City metropolitan region. (A and B) Origins of the viral introductions from other states in Mexico (A) and from other countries (B). States and countries are colored according to the number of introductions

(Continued on next page)

Mexico. There has been recent worsening of the HIV epidemic in this state (12) and, because it is a popular vacation destination, visitors from across Mexico and the world frequent Quintana Roo (located in the Yucatan peninsula). Even more striking was the number of viral introductions from the U.S. state of California, with over 400 inferred events. As in other previously published reports (13–19), these results highlight the importance of the U.S. HIV epidemic to the Mexican HIV epidemic, as well as the close ties of the major metropolitan centers of California to Mexico City and not only to the border state of Baja California. The inferred transmission links with other countries, including Canada, Brazil, and Argentina in the Americas, Spain, France, Germany, Sweden, and the United Kingdom in Europe, and China, Japan, South Korea, and the Philippines in Asia, are also noteworthy. This reflects the role of Mexico City as a global business and tourist hub. Indeed, direct flights now exist between Mexico City and key cities of many of the countries with inferred transmission links (including Los Angeles and San Francisco in California), facilitating easier and faster connections.

On a regional level, our analysis illustrates the complexity of HIV transmission within the Mexico City metropolitan area and also identifies an especially active transmission area involving a few municipalities in the northeast region of the city, i.e., Cuauhtémoc, Iztapalapa, and Gustavo A. Madero. The epidemic in this region has become more important during the past decade and may be a target for focused interventions. Transmissions in this area are dominated by MSM and, unsurprisingly, the Cuauhtémoc borough includes some of better-known gay-friendly neighborhoods in the city. Importantly, Cuauhtémoc and Iztapalapa harbor the two branches of the Condesa clinic. Additionally, Iztapalapa and Gustavo A. Madero are the two most highly populated boroughs of the city (20, 21), while Cuauhtémoc (together with Iztacalco and Benito Juárez) is one of the most densely populated. As shown in our analyses, population size was correlated strongly with the number of migration events to/from each municipality. This observation suggests that interventions for early diagnosis and treatment in these boroughs may have a larger impact on the overall epidemic than interventions randomly dispersed throughout the region. Continued phylodynamic monitoring will allow us to see whether interventions targeted to these areas have an impact on future transmission regionally.

Interestingly, viral migration events between heterosexual persons showed a different phylodynamic and phylogeographic pattern, in comparison to those between MSM. The majority of events involving only heterosexual persons linked municipalities in Mexico City with those in the State of Mexico (77.6% of the interborough transmission events between heterosexual persons, which accounted for 9.63% of all inferred events). Our group and others previously described marked differences in the sociodemographic profiles of MSM and heterosexual populations involved in the HIV epidemic in Mexico City, with the latter showing overall lower socioeconomic and educational status, as well as later presentation to clinical care (22, 23). It is thus not surprising that the phylodynamics and geographic distribution of the heterosexual epidemic are different from those of MSM and will require different types of interventions, deployed in different locations, to optimize prevention efforts in this population.

Our study has a number of limitations that are common to many observational studies. Although completeness of sampling in the metropolitan zone has improved with time (22), completeness of the local network may still be impacted by missing data from the early years of the program. Also, important biases may occur, given that the Condesa clinics attract mainly persons without social security, living close to the two clinic branches, and persons with social security, especially those living in the south of the city, may attend their corresponding clinics and are missed in the present study. Additionally, hard-to-reach populations, including sex workers, persons who inject drugs (PWID), migrants, and hetero-

**FIG 2** Legend (Continued)

originating from the location. Of note, 75% of all introductions originating from the United States were from California. (C) Timing of external introductions (gold) together with sequences sampled in the region (gray). See Materials and Methods for details about Bayesian phylogeographic inferences. The timing of introduction events was inferred from the timing of the most recent common ancestor (tMRCA) of each identified clade of ≥2 sequences.

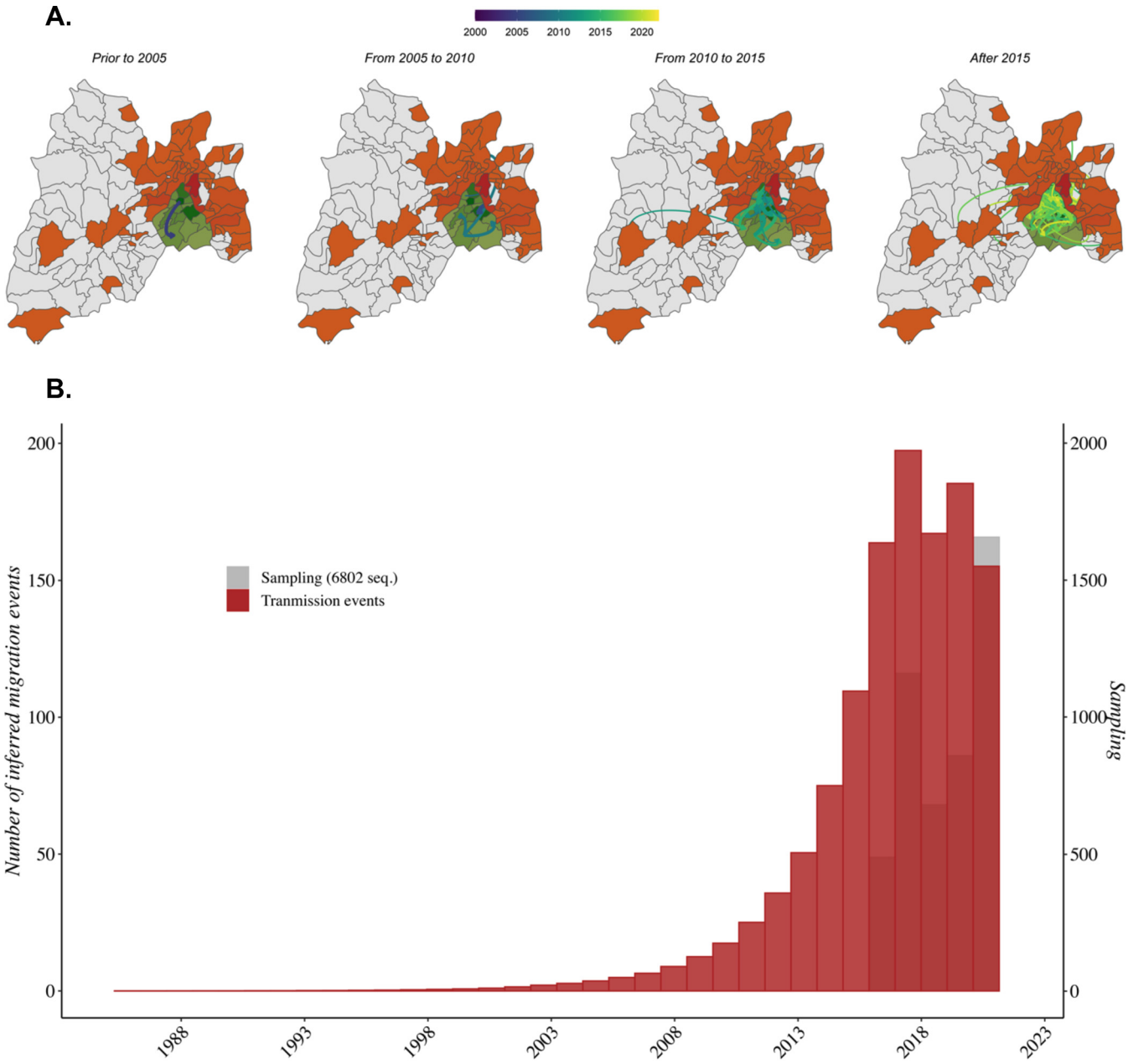


**FIG 3** Inferred migration events between municipalities in the Mexico City metropolitan region over time. (A) Map of the migration events between municipalities. The thickness of the arrows reflects the average number of inferred migration events between municipalities. Municipalities are colored according to the number of sequences in clades (brown tones for Mexico State and green tones for Mexico City). The color of the arrows indicates the timing of the inferred migration event. (B) Timing of inferred migration events (red) together with sequences sampled in the region (gray). See Materials and Methods for details about Bayesian phylogeographic inferences.

sexual cisgender men who are partners of cisgender women with diagnosed infection, who likely do not readily engage in care, may be underrepresented in our study.

While our study illustrates and details the utility of discrete phylogeographic analyses when applied to a comprehensive data set of viral sequences, such spatial investigations are possible only when viral sequences (either as part of local cohorts or available in public databases such as GenBank) are associated with sufficiently precise metadata, particularly the collection date and the sampling location. The availability of precise collection dates for isolates obtained over a sufficient time period enables reliable timing of epidemic events due to the accurate calibration of molecular clock models. Further, phylogeographic inference is possible only when viral genomes are associated with sampling locations.

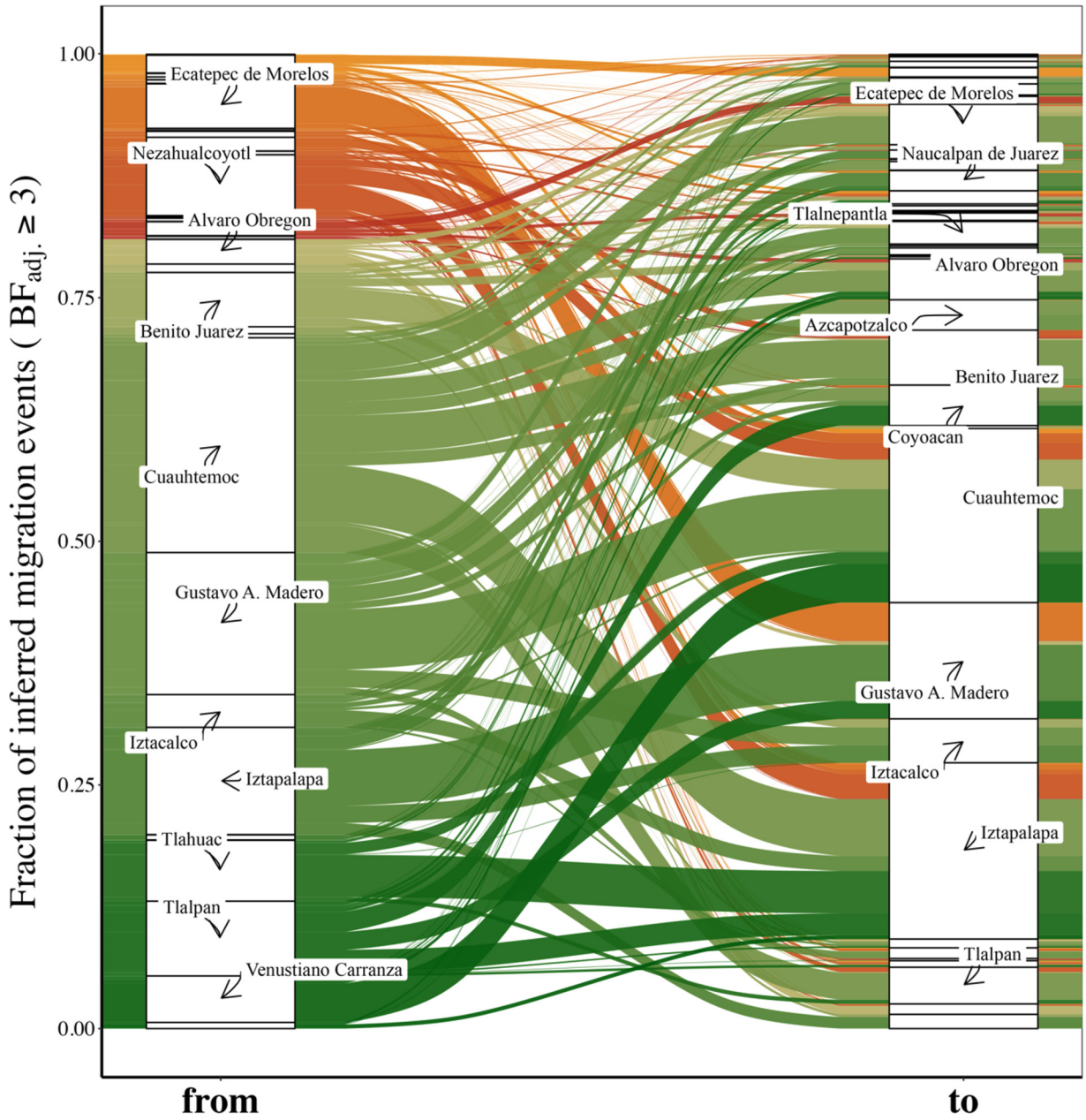


**FIG 4** Proportion of viral migration events from each source location (from) to the recipient location (to). The graph is colored by origin, with Mexico State municipalities shown in brown tones and Mexico City municipalities in green tones. Only locations that account for ≥2% of the events are labeled.

Although our data set is a comprehensive collection of HIV genomes from across Mexico, it is possible that locations remain undersampled. Information on behavioral risk can also be incomplete or incorrect, particularly because cisgender men who are partners of cisgender women living with HIV represent a hard-to-reach population.

For optimal analysis, a sample would be deep, with uniform sampling across demographic and geographic groups. However, sampling in this manner can identify only diagnosed infections, as the number and proportion of undiagnosed individuals remain unknown in these geospatial locations and demographic groups. We illustrate the impact of variation in the numbers of sequences obtained per location in Fig. S2 in the supplemental material, where we show an association between the number of clades observed per municipality and sampling (number of sequence per population size) ($R = 0.61$, $P < 0.0001$). The accuracy and

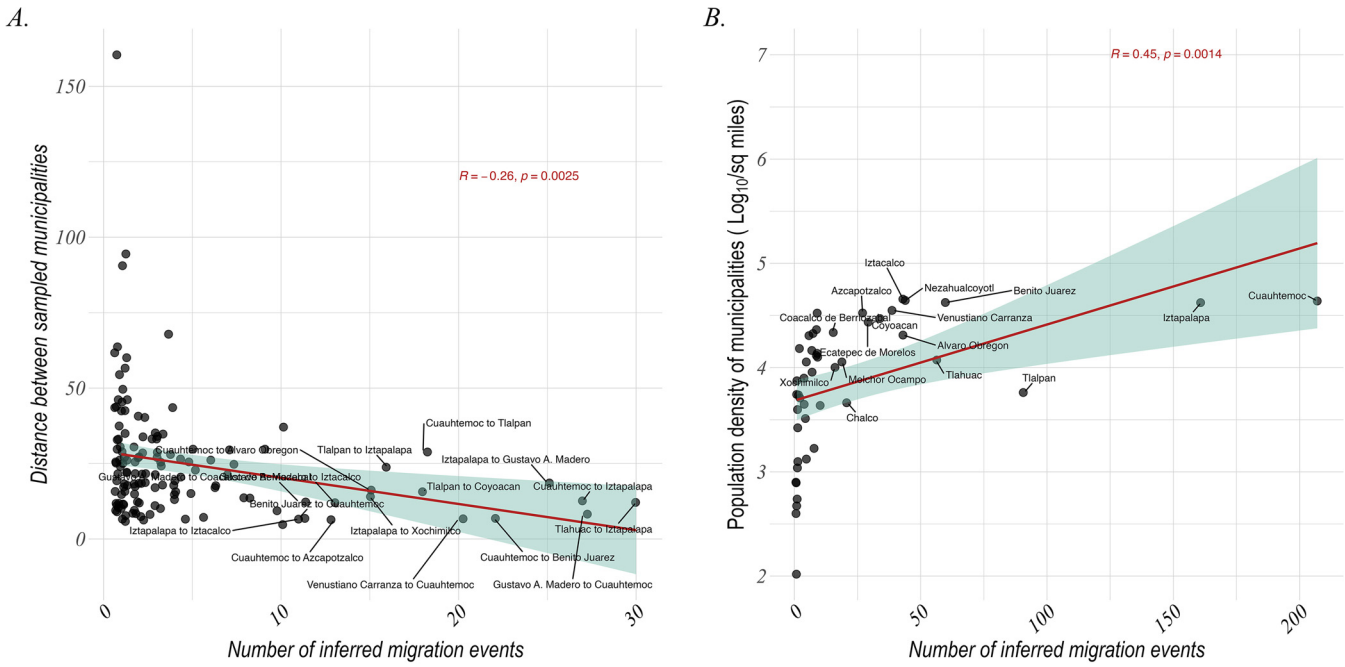


**FIG 5** Association between geospatial distance between municipalities (A) or population density (B) and number of inferred migration events. (A) Association between the number of inferred migration events from/to each municipality and the geospatial distance between municipalities. (B) Association between the total number of inferred migration events (from and to) by municipality and the population density. Only municipalities with ≥10 inferred events are labeled. Population density was defined as the population size estimates by municipality in 2020 obtained from INEGI (53) divided by the area size in square miles.

uniformity of metadata across the sample may also be a limitation, because this could be affected by regional differences in stigma, leading some populations to be more forthcoming than others. While the resolution of phylogeographic analyses always depends on the spatial granularity of available samples, we think that such analyses can still be powerful in elucidating the dispersal history of sampled lineages and highlight the need to improve testing and screening. Finally, we also acknowledge that a fully integrated analysis of all clades, including more than two locations, would be preferable, but the number of sequences available and the number of different sampling locations to consider in the analysis are computationally prohibitive and would require weeks of server runtime in a Bayesian phylogenetic software package such as BEAST.

In conclusion, our study leverages a strong effort to perform HIV molecular surveillance in the Mexico City metropolitan area to infer phylodynamic and phylogeographic characteristics of the local epidemic. We observed a highly complex local epidemic, with strong links to other states in Mexico and several countries around the world, consistent with a large commercial and cultural metropolitan zone. However, we observed a hot spot of HIV transmission, involving three municipalities in the northeast region of Mexico City, that warrants focused interventions. Additionally, we observed marked differences in phylogeographic patterns of heterosexual viral migration versus MSM viral migration, suggesting that the overall epidemic could benefit from interventions targeting venues and locations frequented by persons coming into the city.

## MATERIALS AND METHODS

**Ethics statement.** This study was reviewed and approved by the Ethics Committee and the Ethics in Research Committee of the INER (project codes E02-17 and E02-20) and was conducted according to the principles of the Declaration of Helsinki. All participants provided written informed consent to participate. Additional approval was obtained from the University of California, San Diego, Human Research Protection Program (protocol number 190121).

**Data set.** HIV partial *pol* sequencing was performed for all persons who were recently diagnosed or returning to care after ART interruption at the largest primary HIV clinic in Mexico City and who voluntarily agreed to participate in the study between 2016 and 2021, as described previously (22). Only one sequence (the first sampled) per participant was included in the analysis. Sequences were annotated

with demographic and geographic metadata obtained from a computer-assisted self-administered questionnaire completed at enrollment. Participants' information was kept in a secure server, ensuring confidentiality. Deidentified data sets were used for analysis. The background data set of HIV partial *pol* sequences was obtained from the Los Alamos National Laboratory public HIV sequence database (see below) (24, 25).

**Identification of representative phylogenetic clades.** After sequence curation, a multistep phylogenetic approach was applied to identify putative clades of transmission within the region (across municipalities). We applied the following step-by-step analytical approach. First, transmission clades that best approximate the epidemic dynamics were identified following the method described by Cuypers et al. (10). Briefly, Mexico City and Mexico State sequences were complemented with 9,927 HIV *pol* sequences sampled in other states in Mexico (26, 27), 7,874 sequences sampled in the neighboring U.S. state of California, and 158,661 publicly available HIV *pol* sequences with available location annotation from other countries (one sequence per individual) obtained from the Los Alamos National Laboratory public HIV sequence database (24, 25). For the latter and to reduce the data set size while maintaining an appropriate set of epidemiologically relevant background sequences, we used BLAST (28) to identify the 50 genomes closest to each of the 6,802 genomic sequences sampled in the Mexico City metropolitan region (2, 29), leading to a total of 5,594 background sequences from the rest of the world. These sequences were aligned to the HXB2 *pol* reference sequence (GenBank accession number K03455) (30). Next, phylogenetic trees were inferred using FastTree2 (31) with the GTR+$\Gamma$ substitution model. From these trees, well-supported clades (i.e., Shimodaira-Hasegawa [SH] local support of at least 0.9 [32–34]) including only sequences from Mexico City/State were identified. The purpose of this phylogeographic analysis was to delineate clades corresponding to distinct events of HIV introduction into the Mexico region; we considered an introduction event to be any branch in the phylogenetic tree where the location assigned to a node was Mexico City/State and the location assigned to its parent node in the tree was "other location." Only clades that captured viral migration into the Mexico City/State region were further analyzed.

**Phylogeographic inference.** Phylogeographic inference was performed using the asymmetric discrete phylogeographic model (35, 36) implemented in the BEAST v1.10.5 software package (37). To promote estimation accuracy and precision of the migration rates and the nucleotide substitution rates, the migration and substitution model (GTR+$\Gamma$) was shared across the clades (3, 38). Estimates of the expected number of migration events between all pairs of locations (Markov jumps) were computed through stochastic mapping techniques (39, 40). Specifically, we looked at migration from outside the study region to the study region (viral introductions) and migration within the region (between municipalities).

Simultaneously with reconstructing the geographic migration history, the spread across risk groups was estimated while allowing for sampling uncertainty (41). Briefly, missing risk group information was treated as an ambiguous risk group assignment that could take the value of the other risk groups (that is, the missing data could take the value of heterosexual, MSM, PWID and MSM, or other risks). The starting state of a change in location or risk group is referred to as the "from" location or risk group. Similarly, the location or risk group to which the virus migrated is referred to as the "to" location or risk group.

When sequence data sets lack a clear temporal signal, it is common practice to use independently derived evolutionary rate estimates to specify a suitable prior expectation for the evolutionary rate parameter (42, 43). For subtype B (98.6% of our cohort), the evolutionary rate for *pol* is estimated to be between ~0.001 and ~0.003 substitutions/site/year (44–46). For this reason, we specified a normal distribution as prior on the mean clock rate, with a mean of 0.002 substitutions/site/year and a standard deviation such that the 95% CI ranges from 0.001 to 0.003 substitutions/site/year (44–46). Given that most inferred clades are small and lack temporal signal, we opted to model the rate of evolutionary change in clades with ≥10 taxa with a relaxed clock model (47). To infer the timing of introductions of the smaller clades, the phylogeny was rescaled into units of time with treedater (48), assuming a strict molecular clock with the rate of HIV genome evolution drawn from an externally estimated distribution for HIV-1 subtype B with a mean of 0.002 substitutions/site/year (44–46). To further incorporate uncertainty in the estimated clock rate, molecular clock estimation was replicated 100 times, and the average timing of introduction was inferred. For the same reason, a constant size coalescent tree prior was specified for all clades.

Next, we aimed at interrogating HIV dispersal within the region. To analyze data that are amenable to Bayesian phylogeographic inference, given the statistical uncertainty related to the inference of tree topologies for small clades, we included only clades of size ≥10 sequences ($n$ = 89) originating from at least two distinct municipalities. To identify the subset of migration rates that was most informative to reconstruct the dispersal history, we used a model averaging procedure (Bayesian stochastic search variable selection [BSSVS]) (35), with which it is conveniently possible to obtain Bayes factor (BF) support for all possible migration rates in a single analysis (35). To account for the relative abundances of samples from each location, we measured the $BF_{adj}$ based on a methodology similar to the tipdate randomization test for temporal signal (49, 50). We consider only links with $BF_{adj}$ support of ≥3 for the analyses (51). Markov chain Monte Carlo (MCMC) chains were run long enough to ensure adequate mixing. Maximum clade credibility (MCC) trees were obtained with TreeAnnotator v1.10 (37), and convergence and mixing properties were inspected using Tracer v1.7 (52). The temporal and geospatial distribution of these clades can be visualized in MicroReact (11) (https://microreact.org/project/223ln8u9trx5cqxdaeyfsu) and in Fig. S4 in the supplemental material.

## SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

**SUPPLEMENTAL FILE 1**, PDF file, 2.7 MB.

## ACKNOWLEDGMENTS

This work is supported by grants from the University of California, San Diego, San Diego Center for AIDS Research (SD CFAR) (grant AI036214) (an NIH-funded program that is supported by the following NIH institutes and centers: NIAID, NCI, NHLBI, NIA, NICHD, NIDA, NIDCR, NIDDK, NIGMS, NIMH, NIMHD, FIC, and OAR), Consejo Nacional de Ciencia y Tecnología (CONACyT-PRONAII Virología grant 303079), the Canadian Institutes of Health Research (grants PJT-148621 and PJT-159625), the Research Program on Migration and Health (PIMSA) of the University of California Office of the President (UCOP) (grant S0726506), and the Mexican Government (Programa Presupuestal P016; Anexo 13 del Decreto del Presupuesto de Egresos de la Federación). A.C. is supported by the James B. Pendleton Charitable Trust, SD CFAR, the UCOP (grant R00RG2725), and the NIH NIDA (grant R01DA055491). S.R.M. is funded by SD CFAR and the NIH (grants R24AI106039, R01MH124590, R01AI35992, and R01049644).

We have no financial interest to report.

## REFERENCES

1. WHO. 2021. HIV/AIDS. https://www.who.int/news-room/fact-sheets/detail/hiv-aids.
2. Graf T, Vrancken B, Maletich Junqueira D, de Medeiros RM, Suchard MA, Lemey P, Esteves de Matos Almeida S, Pinto AR. 2015. Contribution of epidemiological predictors in unraveling the phylogeographic history of HIV-1 subtype C in Brazil. J Virol 89:12341–12348. https://doi.org/10.1128/JVI.01681-15.
3. Perez AB, Vrancken B, Chueca N, Aguilera A, Reina G, Garcia-Del Toro M, Vera F, Von Wichman MA, Arenas JI, Tellez F, Pineda JA, Omar M, Bernal E, Rivero-Juarez A, Fernandez-Fuertes E, de la Iglesia A, Pascasio JM, Lemey P, Garcia F, Cuypers L. 2019. Increasing importance of European lineages in seeding the hepatitis C virus subtype 1a epidemic in Spain. Euro Surveill 24:1800227. https://doi.org/10.2807/1560-7917.ES.2019.24.9.1800227.
4. Xia Q, Wertheim JO, Braunstein SL, Misra K, Udeagu CC, Torian LV. 2020. Use of molecular HIV surveillance data and predictive modeling to prioritize persons for transmission-reduction interventions. AIDS 34:459–467. https://doi.org/10.1097/QAD.0000000000002452.
5. Little SJ, Kosakovsky Pond SL, Anderson CM, Young JA, Wertheim JO, Mehta SR, May S, Smith DM. 2014. Using HIV networks to inform real time prevention interventions. PLoS One 9:e98443. https://doi.org/10.1371/journal.pone.0098443.
6. Wertheim JO, Kosakovsky Pond SL, Little SJ, De Gruttola V. 2011. Using HIV transmission networks to investigate community effects in HIV prevention trials. PLoS One 6:e27775. https://doi.org/10.1371/journal.pone.0027775.
7. Vrancken B, Cuypers L, Perez AB, Chueca N, Anton-Basantas J, de la Iglesia A, Fuentes J, Pineda JA, Tellez F, Bernal E, Rincon P, Von Wichman MA, Fuentes A, Vera F, Rivero-Juarez A, Jimenez M, Vandamme AM, Garcia F. 2019. Cross-country migration linked to people who inject drugs challenges the long-term impact of national HCV elimination programmes. J Hepatol 71:1270–1272. https://doi.org/10.1016/j.jhep.2019.08.010.
8. UNAIDS. 2016. Country report Mexico. https://www.unaids.org/sites/default/files/country/documents/MEX_2019_countryreport.pdf. Accessed 28 April 2022.
9. Gobierno de la Ciudad de Mexico. 2021. Respuesta epidemiológica al VIH Sida y al virus de la hepatitis C en la Ciudad de México 2021. https://condesacdmx.mx/index.php/cifras. Accessed 30 April 2022.
10. Cuypers L, Vrancken B, Fabeni L, Marascio N, Cento V, Di Maio VC, Aragri M, Pineda-Pena AC, Schrooten Y, Van Laethem K, Balog D, Foca A, Torti C, Nevens F, Perno CF, Vandamme AM, Ceccherini-Silberstein F. 2017. Implications of hepatitis C virus subtype 1a migration patterns for virus genetic sequencing policies in Italy. BMC Evol Biol 17:70. https://doi.org/10.1186/s12862-017-0913-3.
11. Argimón S, Abudahab K, Goater RJE, Fedosejev A, Bhai J, Glasner C, Feil EJ, Holden MTG, Yeats CA, Grundmann H, Spratt BG, Aanensen DM. 2016. Microreact: visualizing and sharing data for genomic epidemiology and phylogeography. Microb Genom 2:e000093. https://doi.org/10.1099/mgen.0.000093.
12. National Institute of Geography and Statistics, Mexico (INEGI). 2021. https://www.inegi.org.mx/contenidos/saladeprensa/aproposito/2021/EAP_LUCHAVSSIDA_21.pdf. Accessed 15 April 2022.
13. Mehta SR, Wertheim JO, Brouwer KC, Wagner KD, Chaillon A, Strathdee S, Patterson TL, Rangel MG, Vargas M, Murrell B, Garfein R, Little SJ, Smith DM. 2015. HIV transmission networks in the San Diego-Tijuana border region. EBioMedicine 2:1456–1463. https://doi.org/10.1016/j.ebiom.2015.07.024.
14. Mehta SR, Delport W, Brouwer KC, Espitia S, Patterson T, Pond SK, Strathdee SA, Smith DM. 2010. The relatedness of HIV epidemics in the United States-Mexico border region. AIDS Res Hum Retroviruses 26:1273–1277. https://doi.org/10.1089/aid.2010.0021.
15. Zúñiga ML, Brennan J, Scolari R, Strathdee SA. 2008. Barriers to HIV care in the context of cross-border health care utilization among HIV-positive persons living in the California/Baja California US-Mexico border region. J Immigrant Minority Health 10:219–227. https://doi.org/10.1007/s10903-007-9073-7.
16. Strathdee SA, Magis-Rodriguez C, Mays VM, Jimenez R, Patterson TL. 2012. The emerging HIV epidemic on the Mexico-U.S. border: an international case study characterizing the role of epidemiology in surveillance and response. Ann Epidemiol 22:426–438. https://doi.org/10.1016/j.annepidem.2012.04.002.
17. Servin AE, Muñoz FA, Zúñiga ML. 2014. Healthcare provider perspectives on barriers to HIV-care access and utilisation among Latinos living with HIV in the US-Mexico border. Cult Health Sex 16:587–599. https://doi.org/10.1080/13691058.2014.886131.
18. Zhang X, Martinez-Donate AP, Simon NE, Hovell MF, Rangel MG, Magis-Rodriguez C, Sipan CL. 2017. Risk behaviours for HIV infection among travelling Mexican migrants: the Mexico-US border as a contextual risk factor. Glob Public Health 12:65–83. https://doi.org/10.1080/17441692.2016.1142591.
19. Ulibarri MD, Strathdee SA, Patterson TL. 2010. Sexual and drug use behaviors associated with HIV and other sexually transmitted infections among female sex workers in the Mexico-US border region. Curr Opin Psychiatry 23:215–220. https://doi.org/10.1097/YCO.0b013e32833864d5.
20. Mexico City Government. 2020. https://data.consejeria.cdmx.gob.mx/images/leyes/leyes/LEY_ORGANICA_DE_ALCALDIAS_DE_LA_CDMX_5.5.pdf. Accessed 15 April 2022.
21. Wikipedia.com. 2022. https://es.wikipedia.org/wiki/Demarcaciones_territoriales_de_la_Ciudad_de_México. Accessed 15 April 2022.
22. Davila-Conn V, Garcia-Morales C, Matias-Florentino M, Lopez-Ortiz E, Paz-Juarez HE, Beristain-Barreda A, Cardenas-Sandoval M, Tapia-Trejo D, Lopez-Sanchez DM, Becerril-Rodriguez M, Garcia-Esparza P, Macias-Gonzalez I, Iracheta-Hernandez P, Weaver S, Wertheim JO, Reyes-Teran G, Gonzalez-Rodriguez A, Avila-Rios S. 2021. Characteristics and growth of the genetic HIV transmission network of Mexico City during 2020. J Int AIDS Soc 24:e25836. https://doi.org/10.1002/jia2.25836.
23. Bautista-Arredondo S, Servan-Mori E, Beynon F, Gonzalez A, Volkow P. 2015. A tale of two epidemics: gender differences in socio-demographic characteristics and sexual behaviors among HIV positive individuals in Mexico City. Int J Equity Health 14:147. https://doi.org/10.1186/s12939-015-0286-4.
24. Los Alamos National Laboratory. 2022. HIV sequence database. https://www.hiv.lanl.gov/content/sequence/HIV/HIVTools.html. Accessed 24 January 2022.
25. Foley BL, Apetrei C, Hahn B, Mizrachi I, Mullins J, Rambaut A, Wolinsky S, Korber B. 2018. HIV sequence compendium 2018. Los Alamos National Laboratory, Los Alamos, NM.
26. Garcia-Morales C, Tapia-Trejo D, Quiroz-Morales VS, Navarro-Alvarez S, Barrera-Arellano CA, Casillas-Rodriguez J, Romero-Mora KA, Gomez-Palacio-Schjetnan M, Murakami-Ogasawara A, Avila-Rios S, Reyes-Teran G, HIVDR MexNet Group. 2017. HIV pretreatment drug resistance trends

in three geographic areas of Mexico. J Antimicrob Chemother 72:3149–3158. https://doi.org/10.1093/jac/dkx281.
27. Avila-Rios S, Garcia-Morales C, Valenzuela-Lara M, Chaillon A, Tapia-Trejo D, Perez-Garcia M, Lopez-Sanchez DM, Maza-Sanchez L, Del Arenal-Sanchez SJ, Paz-Juarez HE, Quiroz-Morales VS, Mehta SR, Smith DM, Leon-Juarez EA, Magis-Rodriguez C, Reyes-Teran G, HIVDR MexNet Group. 2019. HIV-1 drug resistance before initiation or re-initiation of first-line ART in eight regions of Mexico: a sub-nationally representative survey. J Antimicrob Chemother 74:1044–1055. https://doi.org/10.1093/jac/dky512.
28. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. BMC Bioinformatics 10:421. https://doi.org/10.1186/1471-2105-10-421.
29. Vrancken B, Adachi D, Benedet M, Singh A, Read R, Shafran S, Taylor GD, Simmonds K, Sikora C, Lemey P, Charlton CL, Tang JW. 2017. The multifaceted dynamics of HIV-1 transmission in northern Alberta: a combined analysis of virus genetic and public health data. Infect Genet Evol 52:100–105. https://doi.org/10.1016/j.meegid.2017.04.005.
30. Smith TF, Waterman MS. 1981. Identification of common molecular subsequences. J Mol Biol 147:195–197. https://doi.org/10.1016/0022-2836(81)90087-5.
31. Price MN, Dehal PS, Arkin AP. 2010. FastTree 2: approximately maximum-likelihood trees for large alignments. PLoS One 5:e9490. https://doi.org/10.1371/journal.pone.0009490.
32. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol 59:307–321. https://doi.org/10.1093/sysbio/syq010.
33. Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol 52:696–704. https://doi.org/10.1080/10635150390235520.
34. Shimodaira H, Hasegawa M. 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol Biol Evol 16:1114. https://doi.org/10.1093/oxfordjournals.molbev.a026201.
35. Lemey P, Rambaut A, Drummond AJ, Suchard MA. 2009. Bayesian phylogeography finds its roots. PLoS Comput Biol 5:e1000520. https://doi.org/10.1371/journal.pcbi.1000520.
36. Edwards CJ, Suchard MA, Lemey P, Welch JJ, Barnes I, Fulton TL, Barnett R, O'Connell TC, Coxon P, Monaghan N, Valdiosera CE, Lorenzen ED, Willerslev E, Baryshnikov GF, Rambaut A, Thomas MG, Bradley DG, Shapiro B. 2011. Ancient hybridization and an Irish origin for the modern polar bear matriline. Curr Biol 21:1251–1258. https://doi.org/10.1016/j.cub.2011.05.058.
37. Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A. 2018. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. Virus Evol 4:vey016. https://doi.org/10.1093/ve/vey016.
38. Faria NR, Hodges-Mameletzis I, Silva JC, Rodés B, Erasmus S, Paolucci S, Ruelle J, Pieniazek D, Taveira N, Treviño A, Gonçalves MF, Jallow S, Xu L, Camacho RJ, Soriano V, Goubau P, de Sousa JD, Vandamme A-M, Suchard MA, Lemey P. 2012. Phylogeographical footprint of colonial history in the global dispersal of human immunodeficiency virus type 2 group A. J Gen Virol 93:889–899. https://doi.org/10.1099/vir.0.038638-0.
39. Minin VN, Suchard MA. 2008. Counting labeled transitions in continuous-time Markov models of evolution. J Math Biol 56:391–412. https://doi.org/10.1007/s00285-007-0120-8.
40. Minin VN, Bloomquist EW, Suchard MA. 2008. Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics. Mol Biol Evol 25:1459–1471. https://doi.org/10.1093/molbev/msn090.
41. Scotch M, Tahsin T, Weissenbacher D, O'Connor K, Magge A, Vaiente M, Suchard MA, Gonzalez-Hernandez G. 2019. Incorporating sampling uncertainty in the geospatial assignment of taxa for virus phylogeography. Virus Evol 5:vey043. https://doi.org/10.1093/ve/vey043.
42. Al-Qahtani AA, Baele G, Khalaf N, Suchard MA, Al-Anazi MR, Abdo AA, Sanai FM, Al-Ashgar HI, Khan MQ, Al-Ahdal MN, Lemey P, Vrancken B. 2017. The epidemic dynamics of hepatitis C virus subtypes 4a and 4d in Saudi Arabia. Sci Rep 7:44947. https://doi.org/10.1038/srep44947.
43. Zhang Y, Vrancken B, Feng Y, Dellicour S, Yang Q, Yang W, Zhang Y, Dong L, Pybus OG, Zhang H, Tian H. 2017. Cross-border spread, lineage displacement and evolutionary rate estimation of rabies virus in Yunnan Province, China. Virol J 14:102. https://doi.org/10.1186/s12985-017-0769-6.
44. Abecasis AB, Vandamme AM, Lemey P. 2009. Quantifying differences in the tempo of human immunodeficiency virus type 1 subtype evolution. J Virol 83:12917–12924. https://doi.org/10.1128/JVI.01022-09.
45. Patino-Galindo JA, Gonzalez-Candelas F. 2017. The substitution rate of HIV-1 subtypes: a genomic approach. Virus Evol 3:vex029. https://doi.org/10.1093/ve/vex029.
46. Vrancken B, Baele G, Vandamme AM, van Laethem K, Suchard MA, Lemey P. 2015. Disentangling the impact of within-host evolution and transmission dynamics on the tempo of HIV-1 evolution. AIDS 29:1549–1556. https://doi.org/10.1097/QAD.0000000000000731.
47. Drummond AJ, Ho SYW, Phillips MJ, Rambaut A. 2006. Relaxed phylogenetics and dating with confidence. PLoS Biol 4:e88. https://doi.org/10.1371/journal.pbio.0040088.
48. Volz EM, Frost SDW. 2017. Scalable relaxed clock phylogenetic dating. Virus Evol 3:vex025. https://doi.org/10.1093/ve/vex025.
49. Vrancken B, Mehta SR, Ávila-Ríos S, García-Morales C, Tapia-Trejo D, Reyes-Terán G, Navarro-Álvarez S, Little SJ, Hoenigl M, Pines HA, Patterson T, Strathdee SA, Smith DM, Dellicour S, Chaillon A. 2021. Dynamics and dispersal of local human immunodeficiency virus epidemics within San Diego and across the San Diego-Tijuana border. Clin Infect Dis 73:e2018–e2025. https://doi.org/10.1093/cid/ciaa1588.
50. Trovão NS, Baele G, Vrancken B, Bielejec F, Suchard MA, Fargette D, Lemey P. 2015. Host ecology determines the dispersal patterns of a plant virus. Virus Evol 1:vev016. https://doi.org/10.1093/ve/vev016.
51. Kass RE, Raftery AE. 1995. Bayes factors. J Am Stat Assoc 90:773–795. https://doi.org/10.1080/01621459.1995.10476572.
52. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. 2018. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Syst Biol 67:901–904. https://doi.org/10.1093/sysbio/syy032.
53. National Institute of Geography and Statistics, Mexico (INEGI). 2020. https://www.inegi.org.mx/app/scitel/Default?ev=9. Accessed 15 April 2022.